AD-A094 409 TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS F/G 12/1
TIME SERIES FORECASTING BY ARARMA MODELS.(U)
JUN 80 E PARZEN N00014-78-C-0599
UNCLASSIFIED TR-N-14 NL

1 OF 1
AD-A094409

END
DATE FILMED
2-81
DTIC

LEVEL

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843

Institute of Statistics

TIME SERIES FORECASTING
BY ARARMA MODELS

by Emanuel Parzen
Institute of Statistics
Texas A&M University

Technical Report No. N-14
Jun[illegible]80

TR-N-14

Texas A & M Research Foundation
Project No. 3838

"Multiple Time Series Modeling and Time Series Theoretic Statistical Methods"

Sponsored by the Office of Naval Research

Professor Emanuel Parzen, Principal Investigator

Approved for public release; distribution unlimited.

AD A094409

DDC FILE COPY

DTIC ELECTE FEB 2 1981 S C

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: Technical Report N-14	2. GOVT ACCESSION NO.: AD-A094409	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Time Series Forecasting by ARARMA Models		5. TYPE OF REPORT & PERIOD COVERED: Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Emanuel Parzen		8. CONTRACT OR GRANT NUMBER(s): N00014-78-C-0599
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Texas A&M University, Institute of Statistics, College Station, TX 77843		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Arlington, Virginia 22217, Code 436		12. REPORT DATE: June, 1980
		13. NUMBER OF PAGES: 49
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report): Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report): NA		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): Forecasting, Stationary Time Series, Non-stationary Time Series, Autoregression, Prediction Horizon, ARIMA Models, ARARMA Models.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number): The relation between various basic problems of time series analysis is presented ~~in Section 2~~. Exponential smoothing methods are developed from the point of view of prediction theory ~~(section 2)~~ and extended ~~(section 3)~~. ARARMA models are introduced ~~(section 4)~~. Methods of ARARMA model fitting are outlined ~~(sections 5,6)~~. Since "the proof of the pudding is in the eating," the methods proposed are illustrated ~~(sections 7-10)~~ using some classic examples of time series, including international airline passengers, Makridakis metals series and sunspots.		

DD FORM 1473, 1 JAN 73 — EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

81 1 30 038

TIME SERIES FORECASTING BY ARARMA MODELS

by Emanuel Parzen
Institute of Statistics
Texas A&M University

Abstract

Methods of time series forecasting are proposed which *can be applied automatically*. However, they are not rote formulas, since they are based on a flexible philosophy which provides several models for consideration and diverse diagnostics for qualitatively and quantitatively checking the fit of a model. The models considered are called ARARMA models because the model computed adaptively for a time series is based on sophisticated time series analysis of ARMA schemes (a short memory model) fitted to residuals of simple extrapolation (a long memory model obtained by parsimonious "best lag" non-stationary autoregression).

A major problem of time series forecasting is whether long range forecasting and short range forecasting require different methods to obtain satisfactory forecasts. This paper describes iterated (or tandem) models which provide qualitative diagnostics as to the possibility of long range forecasts (by diagnosing whether the time series is long memory). Both long range and short range forecasts are provided by an ARARMA model.

* This research was supported by the Office of Naval Research (Contract N00014-78-C-0599).

The relation between various basic problems of time series analysis is presented in section 1. Exponential smoothing methods are developed from the point of view of prediction theory (section 2) and extended (section 3). ARARMA models are introduced (section 4). Methods of ARARMA model fitting are outlined (sections 5, 6). Since "the proof of the pudding is in the eating," the methods proposed are illustrated (sections 7-10) using some classic examples of time series, including international airline passengers, Makridakis metals series and sunspots.

1. Some Basic Problems of Time Series Analysis

Time series analysis is concerned with learning about a time series $Y(t)$, $t = 0, \pm 1, \ldots$ from an observed sample $\{Y(t),\ t = 1, 2, \ldots, T\}$. One may distinguish four main problems: forecasting, spectral analysis, parameter estimation, and model identification.

Forecasting is concerned with forming, for horizons $h = 1, 2, \ldots$, predictions of $Y(t+h)$ given $Y(t), Y(t-1), \ldots, Y(t-m+1)$, called the memory m predictor of the time series at time t. The predictor is denoted $Y^{\mu,m}(t+h|t)$. The prediction error is

$$Y^{\nu,m}(t+h|t) = Y(t+h) - Y^{\mu,m}(t+h\ t) .$$

When $m = \infty$, the superscript m is omitted.

The normalized mean square prediction error of infinite memory prediction h steps ahead is denoted

$$\sigma^2_{h,\infty} = E[|Y^{\nu}(t+h|t)|^2] \div E[|Y(t)|^2] .$$

Parzen (1981) proposes the prediction variance horizon function

$$PVH(h) = 1 - \sigma^2_{h,\infty} \quad , \quad h = 1, 2, \ldots,$$

as a very useful diagnostic tool for model identification.

By spectral analysis of a discrete parameter time series $Y(t)$ one means determination of the properties of the function $\dot{Y}(\lambda)$ in the following spectral representation of the time series [see Parzen (1981)]:

$$Y(t) = \int_{-0.5}^{0.5} \exp(2\pi i\lambda t)\ dY(\lambda) .$$

When the time series $Y(t)$ has zero means and is stationary in the sense that there is a covariance function $R(v)$, $v = 0, \pm 1, \ldots$, satisfying for all t

$$R(v) = E[Y(t)\ Y(t+v)] .$$

then $Y(\lambda)$ is a stochastic integrand satisfying

$$E|dY(\lambda)|^2 = R(0)f(\lambda)\ d\lambda = R(0)\ dF(\lambda)$$

where $f(\lambda)$ and $F(\lambda)$ are spectral density and spectral distribution functions defined as follows. Call

$$\rho(v) = R(v)/R(0) = \mathrm{Corr}[Y(t), Y(t+v)]$$

the correlation function. If $\Sigma|\rho(v)| < \infty$, define

$$f(\lambda) = \sum_{v=-\infty}^{\infty} e^{-2\pi i\lambda v}\ \rho(v) \ , \quad F(\lambda) = 2\int_0^{\lambda} f(\lambda')\ d\lambda' .$$

Parameter estimation is concerned with the efficient estimation (using criteria of confirmatory statistical inference such as maximum likelihood) of the parameters of probability models for time series. An important model is a whitening filter representation

$$\sum_{j=0}^{p} \alpha_j\ Y(t-j) = \sum_{k=0}^{q} \beta_k\ \varepsilon(t-k)$$

where $\varepsilon(t)$ is white noise, that is, zero mean uncorrelated Gaussian random variables with variance $\sigma^2 = E[|\varepsilon(t)|^2]$. The whitening filter is called AR(p) if q = 0; MA(q) if p = 0; and ARMA(p,q) otherwise. The parameters α_j and β_k are usually constrained to satisfy:

$$g(z) = \sum_{j=0}^{p} \alpha_j z^j , \quad h(t) = \sum_{k=0}^{q} \beta_k z^k$$

have all their roots in the complex z-plane outside the unit circle. A parameter estimation problem is to estimate $\sigma^2, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$. A model identification problem is to estimate the orders p and q. A different type of model identification problem is concerned with model type, and in particular whether the time series can be assumed to be ARMA, or is even stationary.

Model Identification is the fundamental problem of time series analysis, because its solution implies solutions to the other problems mentioned. It is also the most difficult problem to define. Akaike (1976) proposes that "A procedure of model identification may be defined as a procedure of finding a model which fits best to a set of observed data with respect to a criterion." As the criterion of fit of a model one can take the quality of its parameter estimates, spectral estimates, and forecasts. Parzen (1979)-(1981) proposes that there is an equivalence between the four basic questions of time series analysis, and they must be solved simultaneously, and proposes various iterated time series models that can be fitted to data "semi-automatically" on the basis of their

"signatures" and that can be interpreted to solve the various basic problems of time series analysis. This paper introduces a simple form of iterated model, called ARARMA models, which are particularly suitable for forecasting. We believe that the forecasting procedure proposed in this paper satisfies the advice of Einstein: "everything should be made as simple as possible but not simpler."

ARARMA models can be regarded as an extension of ARIMA models introduced by Box and Jenkins (1976). ARIMA models treat as a stationary ARMA model either Y(t) or a time series $\tilde{Y}(t)$ obtained by "differencing" operations of the form

$$\tilde{Y}(t) = \nabla Y(t) = Y(t+1)-Y(t) \text{ or } \tilde{Y}(t) = \nabla_\tau Y(t) = Y(t+\tau)-Y(t)$$

where τ is the period of the seasonal effect. The model identification problem consists of estimating the orders p and q, and the differencing operators required.

By introducing the lag operator L, LY(t) = Y(t-1), the ARIMA (p,d,q) model is expressed

$$g(L)\ \nabla^d\ Y(t) = h(L)\ \varepsilon(t) .$$

A parametrization is called parsimonious if it has as few parameters as possible. In the analysis of seasonal time series with period τ, parsimony is sought through "multiplicative" or "factored" models of the form

$$g(L)\ G(L^\tau)\ \nabla^d\ \nabla_\tau^D\ Y(t) = h(L)\ H(L^\tau)\varepsilon(t)$$

where G and H are polynomials of degree P and Q respectively. The model is denoted ARIMA$(p,d,q) \times (P,D,Q)_\tau$.

Some objections to the ARIMA model fitting approach are: (1) the determination of the seasonal periods τ is completely subjective (for example, in monthly economic data, both quarterly (τ=3) and annual (τ=12) "periods" may be present); (2) the ARIMA model might suffice for prediction, but does not provide a conceptual framework for a decomposition of the time series into trend, seasonal components, and stationary residual (Parzen (1980)); (3) the ARIMA modeling strategy does not take account of the fact that the prediction effectiveness of an ARIMA model may be mainly determined by the differencing operation $\nabla^d \nabla_\tau^D$, and not by the ARMA part of the model.

2. Prediction Theory and Exponential Smoothing

Our time series analysis approach to forecasting is based upon the following powerful notation for forecasts:

$$Y^\mu(t+1) = \text{predictor of } Y(t+1) \text{ given } Y(t),\ Y(t-1),\ \ldots$$

$$= E[Y(t+1)|Y(t),\ Y(t-1),\ \ldots] \ ;$$

$$Y^\nu(t+1) = \text{prediction error of } Y(t+1) \text{ given } Y(t),\ Y(t-1),\ \ldots$$

$$= Y(t+1) - Y^\mu(t+1|t) \ ;$$

$$Y^\mu(t+h|t) = \text{predictor of } Y(t+h) \text{ given } Y(t),\ Y(t-1),\ \ldots$$

$$= E[Y(t+h)|Y(t),\ Y(t-1),\ \ldots] \ ;$$

$$Y^\nu(t+h|t) = \text{prediction error of } Y(t+h) \text{ given } Y(t),\ Y(t),\ \ldots$$

$$= Y(t+h) - Y^\mu(t+h|t) \ .$$

$$\sigma^2_{h,\infty} = E[|Y^\nu(t+h|t)|^2] \div E[|Y(t)|^2]$$

$$= \text{normalized mean square prediction error}$$

If a time series $Y(t)$, $t = 0, \pm 1, \ldots$ is a stationary time series then the solution to the problem of forecasting future values $Y(t+h)$, $h = 1, 2, \ldots$ from $Y(t)$, $Y(t-1)$, ... may be expressed:

$$Y(t+h) = Y^\mu(t+h|t) + Y^\nu(t+h|t)$$

where $Y^\mu(t+h|t)$ is the predictor and $Y^\nu(t+h|t)$ *is the prediction error.*

A general formula for the predictor can be given in terms of the AR(∞), or infinite order autoregressive, representation:

$$Y(t) + \alpha_1 Y(t-1) + \ldots = \varepsilon(t)$$

where $\varepsilon(t)$ is white noise. Then

$$Y(t+h) + \alpha_1 Y(t+h-1) + \ldots = \varepsilon(t+h) \ ,$$

$$Y^\mu(t+h|t) + \alpha_1 Y^\mu(t+h-1|t) + \ldots = 0 \ .$$

Note that $Y^\mu(t+j|t) = Y(t+j)$ for $j \le 0$.

A general formula for mean square prediction error can be expressed in terms of the MA(∞) , or infinite order moving average, representation:

$$Y(t) = \varepsilon(t) + \beta_1 \varepsilon(t-1) + \ldots$$

$$Y(t+h) = \varepsilon(t+h) + \beta_1 \varepsilon(t+h-1) + \ldots + \beta_{h-1}\varepsilon(t+1) + \ldots$$

$$Y^{\nu}(t+h|t) = \varepsilon(t+h) + \beta_1 \varepsilon(t+h-1) + \ldots + \beta_{h-1}\, \varepsilon(t+1)$$

$$\sigma^2_{h,\infty} = \sigma^2_{\infty} \{1 + \beta_1^2 + \ldots + \beta_{h-1}^2\}$$

where $\sigma^2_{\infty} = E[|\varepsilon(t)|^2] \div E[|Y(t)|^2]$.

Given observations up to time t, the future values of the time series are predicted by $Y^{\mu}(t+h|t)$, $h = 1, 2, \ldots$. Consequently the aim of the theory of forecasting is algorithms for computing $Y^{\mu}(t+h|t)$ and $\sigma^2_{h,\infty}$. The approach we recommend for developing such algorithms is best introduced by a discussion of exponential smoothing.

A possible predictor of Y(t+1) given past values is a <u>simple moving average</u>

$$Y^{\mu}(t+1) = \frac{1}{N} \{Y(t) + Y(t-1) + \ldots + Y(t-N+1)\} \quad .$$

To compute it at successive times t one could use the <u>recursive formula</u>

$$Y^{\mu}(t+1) = Y^{\mu}(t) + \frac{1}{N} \{Y(t) - Y(t-N)\} \ .$$

In this formula replace 1/N by a constant α in $0 < \alpha < 1$, and replace Y(t-N) by $Y^{\mu}(t)$; one obtains the recursive smoothing formula called <u>exponential smoothing</u>:

$$\begin{aligned} Y^{\mu}(t+1) &= Y^{\mu}(t) + \alpha\, \{Y(t) - Y^{\mu}(t)\} \\ &= \alpha Y(t) + (1-\alpha)\, Y^{\mu}(t) \end{aligned}$$

A major problem with exponential smoothing is how to choose α; if $\alpha = 1$, $Y^{\mu}(t)$ follows Y(t) very closely, so to smooth one must choose α close to 0, say $\alpha = .1$ or $\alpha = .2$.

To obtain insight into how to estimate α from data, one must define the time series model for which exponential smoothing provides optimal predictors.

To obtain the model implied by a predictor, one must pass from a formula for $Y^{\mu}(t)$ to a formula for Y(t). Now replace $Y^{\mu}(t)$ by $Y(t) - Y^{\nu}(t)$ to obtain

$$Y(t+1) - Y^{\nu}(t+1) = Y(t) - Y^{\nu}(t) + \alpha Y^{\nu}(t) \ ,$$

$$Y(t+1) - Y(t) = Y^{\nu}(t+1) - (1-\alpha)\, Y^{\nu}(t) \ .$$

Let $\beta = 1-\alpha$, $\tilde{Y}(t)$ be the first difference of Y(t): $\tilde{Y}(t+1) = Y(t+1) - Y(t)$, and $\varepsilon(t) = Y^{\nu}(t)$. Then $\tilde{Y}(t)$ satisfies the model

$$\tilde{Y}(t+1) = \varepsilon(t+1) - \beta\varepsilon(t)$$

which is an MA(1), moving average of order 1, if $\varepsilon(t)$ is white noise with variance σ^2. When $Y^{\mu}(t)$ is the infinite memory one step ahead predictor of a stationary Gaussian time series, the innovations $Y^{\nu}(t)$ are white noise. Consequently to estimate α from the data, one estimates the parameter of a MA(1) model for $\tilde{Y}(t)$.

3. <u>Extensions of Exponential Smoothing and Random Walk</u>

A time series Y(t) is said to obey a random walk if

$$Y(t+1) - Y(t) = \epsilon(t+1) .$$

where $\epsilon(t)$ is white noise. Then

$$Y^{\mu}(t+1) = Y(t) \qquad (1)$$

Indeed

$$Y^{\mu}(t+h|t) = Y(t) . \qquad (2)$$

To prove (1) and (2), write

$$Y(t+h) = Y(t) + \epsilon(t+1) + \dots + \epsilon(t+h) ,$$

and take the conditional expectation with respect to $Y(t), Y(t-1), \dots$.

The prediction formulas (1) and (2) are called <u>naive prediction</u>: they are optimal for a time series obeying a random walk. Exponential smoothing is an extension of naive prediction. These two prediction methods correspond to assumptions concerning the time series

$$\tilde{Y}(t+1) = Y(t+1) - Y(t)$$

Let us assume that the original time series $Y(t)$ is non-stationary, but $\tilde{Y}(t)$ is stationary. Then one can model $\tilde{Y}(t)$ and form predictors of it from its own past, denoted $\tilde{Y}^{\mu}(t+1)$ and $\tilde{Y}^{\mu}(t+h|t)$. Since

$$Y(t+h) - Y(t+h-1) = \tilde{Y}(t) ,$$

we obtain the following <u>extensions of exponential smoothing</u>:

$$Y^{\mu}(t+1) = Y(t) + \tilde{Y}^{\mu}(t+1) ,$$

$$Y^{\mu}(t+h|t) = Y^{\mu}(t+h-1|t) + \tilde{Y}^{\mu}(t+h|t) .$$

When $\tilde{Y}(t) = \epsilon(t)$, $\tilde{Y}^{\mu}(t+1) = 0$. When $\tilde{Y}(t+1) = \epsilon(t+1) - \beta\epsilon(t)$, $\tilde{Y}^{\mu}(t+1) = -\beta\epsilon(t)$. Consequently $\tilde{Y}^{\nu}(t+1) = \epsilon(t+1) = Y^{\nu}(t+1)$. This result illustrates a fact which is <u>often</u> true:

$$\tilde{Y}^{\nu}(t) = Y^{\nu}(t) ;$$

in words, <u>the innovations of $Y(t)$ and of $\tilde{Y}(t)$ coincide</u>. We are not restricted to modeling $\tilde{Y}(t)$ as a MA(1); perhaps a more general model is an AR(p):

$$\tilde{Y}(t) + \alpha(1) \tilde{Y}(t-1) + \dots + \alpha(p) \tilde{Y}(t-p) = \tilde{\epsilon}(t) .$$

Then

$$-\tilde{Y}^{\mu}(t+1) = \alpha(1) \tilde{Y}(t) + \dots + \alpha(p) \tilde{Y}(t-p+1)$$

4. <u>ARARMA Time Series Models for Forecasting</u>

A natural extension of the exponential smoothing prediction strategy involves assuming a general stationary ARMA(p,q) model for $\tilde{Y}(t)$, and a general non-stationary autoregressive operator for $\tilde{Y}(t)$. Thus define

$$\tilde{Y}(t) = Y(t) - \phi_1 Y(t-1) - \dots - \phi_r Y(t-r) ,$$

$$\sum_{j=1}^{p} \alpha_j (\tilde{Y}(t-j) - \tilde{\mu}) = \sum_{k=1}^{q} \beta_k \epsilon(t-k) ,$$

where $\varepsilon(t)$ is white noise. This model is of the form

$$\text{Long Memory } Y \longrightarrow \boxed{G_r} \longrightarrow \text{Short Memory } \tilde{Y} \longrightarrow \boxed{g_\infty} \longrightarrow \text{White Noise } \varepsilon$$

emphasized by Parzen (1979), (1980). It includes the ARIMA models of Box and Jenkins (19¯6) as a special case. The model could be called an ARARMA model, to indicate that Y is defined by an autoregressive operator, rather than differencing.

The transfer functions

$$g_p(z) = \sum_{j=0}^{p} \alpha_j z^j , \quad h_q(z) = \sum_{k=0}^{q} \beta_k z^k$$

are assumed to have all their roots outside the unit circle in the complex z-domain. However

$$G_r(z) = 1 - \sum_{j=1}^{r} \phi_j z^j ,$$

is not assumed to obey this stationarity constraint.

Given the foregoing model, one can generate forecasts by recursive formulas such as

$$Y^\mu(t+h|t) = \sum_{j=1}^{r} \phi_j Y^\mu(t+h-j|t) + \tilde{Y}^\mu(t+h|t) ,$$

$$\sum_{j=0}^{p} \alpha_j \{\tilde{Y}^\mu(t+h-j|t) - \bar{\tilde{Y}}\} = \sum_{k=0}^{q} \beta_k \varepsilon^\mu(t+h-k|t).$$

In this paper the model adopted for a time series is usually of the form

$$\tilde{Y}(t) = Y(t) - \phi(\tau) Y(t-\tau) ,$$

$$\tilde{Y}(t) + \alpha_1 \tilde{Y}(t-1) + \dots + \alpha_m \tilde{Y}(t-m) = \varepsilon(t) .$$

From the equation

$$Y(t+h) = \phi(\tau) Y(t-\tau+h) + \tilde{Y}(t+h)$$

one obtains, by conditioning with respect to $Y(t)$, $Y(t-1)$, ...

$$Y^\mu(t+h|t) = \phi(\tau) Y^\mu(t-\tau+h|t) + \tilde{Y}^\mu(t+h|t) .$$

To obtain a formula for forecasts of $\tilde{Y}$ when we have fitted an AR(m) to $\tilde{Y}$, write

$$\tilde{Y}(t+h) + \alpha_1 \tilde{Y}(t+h-1) + \dots + \alpha_m \tilde{Y}(t+h-m) = \varepsilon(t+h) ,$$

$$\tilde{Y}^\mu(t+h|t) + \alpha_1 Y^\mu(t+h-1|t) + \dots + \alpha_m \tilde{Y}^\mu(t+h-m|t) = 0 .$$

One computes $\tilde{Y}^\mu(t+h|t)$ recursively for $h = 1, 2, \dots$, using the fact that $\tilde{Y}^\mu(t+j|t) = \tilde{Y}(t+j)$ if $j \le 0$. For example,

$$-\tilde{Y}^\mu(t+1|t) = \alpha_1 \tilde{Y}(t) + \dots + \alpha_m \tilde{Y}(t-m+1) .$$

Finally, one computes $Y^\mu(t+h|t)$ recursively for $h = 1, 2, \dots$ using the fact that $Y^\mu(t+j|t) = Y(t+j)$ if $j \le 0$.

For large values of h, one expects $\tilde{Y}^\mu(t+h|t) \doteq 0$. Then $Y^\mu(t+h|t) = \phi(\tau) Y^\mu(t+h-\tau|t)$. When $\phi(\tau) \ge 1$, this does not damp down to zero, and provides long range forecasts which essentially extrapolate the time series.

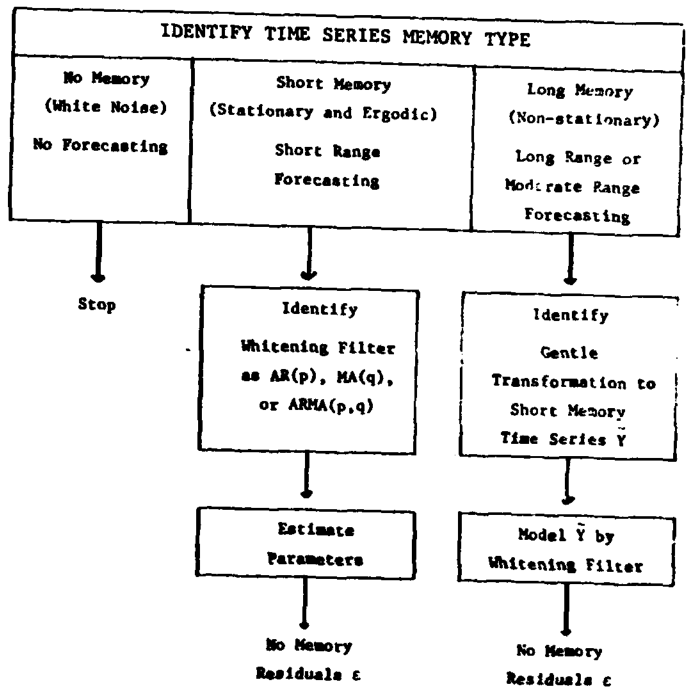
Identifying the whitening filter, and estimating the parameters, of the short memory time series $\tilde{Y}$ can be performed rigorously. Finding the definition of $\tilde{Y}$, or equivalently the best transformation of the original time series Y to a short memory time series $\tilde{Y}$, is an art. However we propose an automatic procedure of the following general form.

To check if $Y(t)$ is long memory and to find a transformation of $Y(t)$ to a short memory time series we consider transformations of the form

$$\tilde{Y}(t) = Y(t) - \phi(\tau)\, Y(t-\tau)$$

where τ is a lag (called a "best lag" and representing a period in the time series) which is determined adaptively by a procedure called "nonstationary best lag autoregression." If $\tilde{Y}(t)$ is short memory, we usually model it for forecasting purposes by an $AR(\hat{m})$ scheme whose order $\hat{m}$ is chosen by an order determining criterion (CAT). In addition, we determine the best subset ARMA scheme for $\tilde{Y}(t)$.

Our approach to univariate time series modeling, forecasting and spectral estimation is based on the concept that to identify a model, or "overall whitening filter", for a time series, one should determine its *model memory* type, and identify an iterative or *tandem* model for the time series:



5. Stationary Analysis of an Empirical Time Series

A confirmatory theory of statistical inference is available only for short memory time series (which are stationary and ergodic). The modeling of a short memory time series by a whitening filter can be regarded as a science, and it can be made semi-automatic. Given a sample of short memory stationary time series $\tilde{Y}(t)$, our modeling procedure in the time domain is to compute approximating autoregressive schemes.

To analyze a time series sample $Y(t)$, $t = 1, \ldots, T$, as a stationary time series, our program ARSPID (Autoregressive Spectral Identification) proceeds as follows.

Step I. Basic Spectral Statistics. Compute using FFT (i) the sample spectral density $\tilde{f}(\lambda)$, (ii) sample correlation function $\tilde{\rho}(v)$, and (iii) sample distribution function $\tilde{F}(\lambda)$.

Step II. Toeplitz Normal Equations Recursion. Solve Yule-Walker equations in $\hat{\rho}(v)$

$$\hat{\rho}(v)+\hat{\alpha}_m(1)\hat{\rho}(v-1)+\ldots+\hat{\alpha}_m(m)\hat{\rho}(v-m) = 0 , \quad v = 1, \ldots, m ,$$

$$\hat{\sigma}_m^2 = 1 + \hat{\alpha}_m(1)\hat{\rho}(1) + \ldots + \hat{\alpha}_m(m)\hat{\rho}(m) .$$

for successive orders $m = 1, 2, \ldots$ using fast recursive algorithms. Optionally one can form autoregressive spectral estimators, defined in terms of the transfer function $g_m(z) = 1+\hat{\alpha}_m(1)z+\ldots+\hat{\alpha}_m(m)z^m$ by

$$\hat{f}_m(\lambda) = \hat{\sigma}_m^2 \, |\hat{g}_m(e^{2\pi i\lambda})|^{-2} , \quad 0 \leq \lambda \leq 0.5 .$$

The autoregressive spectral distribution function estimator of

order m is defined by

$$\hat{F}_m(\lambda) = \int_0^\lambda \hat{f}_m(\lambda')\, d\lambda' , \quad 0 \leq \lambda \leq 0.5 .$$

Step III. Order Determination. Form criteria functions whose absolute minimum and relative minimum are used to determine orders m of autoregressive estimators to be considered as "optimal" Akaike information criterion is [see Akaike (1979), Hannan and Quinn (1979)]

$$AIC(m) = \log \hat{\sigma}_m^2 + \frac{2m}{T}$$

Parzen CAT (Criterion Autoregressive Transfer Function), a measure of the overall mean square relative error of g_m as an estimator of g_∞, (the AR(∞) transfer function), is defined in terms of

$$\hat{\bar{\sigma}}_j^2 = \frac{T}{T-j} \hat{\sigma}_j^2 ,$$

an unbiased estimator of σ_j^2, by [see Parzen (1974), (1977)]

$$CAT(m) = \frac{1}{T} \sum_{j=1}^{m} \hat{\bar{\sigma}}_j^{-2} - \hat{\bar{\sigma}}_m^{-2} .$$

To increase the probability of correctly detecting white noise, one defines

$$CAT(0) = -(1+\frac{1}{T}) , \quad AIC(0) = -\frac{1}{T} .$$

In practice, identical conclusions are usually obtained from AIC and CAT. The best order $\hat{m}$ is defined as the order at

which the criterion function is minimized (if $\hat{m} = 0$, the time series is considered to be white noise or no memory). The second best order $\hat{m}(2)$ is defined as the order at which the smallest relative minimum occurs which is not a global minimum.

The orders selected by the minimum of an AR order criterion function are only hypotheses the ultimate test of which is how well their autoregressive spectral distribution function fits the raw spectral distribution function. Thus one examines

$$\hat{F}(u)-\tilde{F}(u) = \int_0^u \{\sigma_m^2 |g_m(e^{2\pi i\lambda})|^{-2} - \tilde{f}(\lambda)\} d\lambda = \int_0^u \{\hat{f}(\lambda)-\tilde{f}(\lambda)\} d\lambda .$$

In addition, one should examine the residuals of the autoregressive filter for whiteness which is equivalent to examining

$$\int_0^u \{\frac{1}{\sigma_m^2} |g_m(e^{2\pi i\lambda})|^2 \tilde{f}(\lambda) - 1\} d\lambda = \int_0^u \frac{\tilde{f}(\lambda)-\hat{f}(\lambda)}{\hat{f}(\lambda)} d\lambda.$$

STEP IV. Prediction Variance Horizon, Model Memory Identification. A new tool for model identification is the prediction variance horizon function PVH(h), $h = 1, 2, \ldots$ defined in terms of the normalized mean square prediction error of infinite memory prediction h steps ahead: $\sigma^2_{h,\infty} = E[|Y^{\nu}(t+h|t)^2] \div E[Y^2(t)]$.

A formula for $\sigma^2_{h,\infty}$ is obtained by introducing the MA(∞), infinite moving average, representation of Y(t)

$$Y(t) = \epsilon(t) + \beta_1 \epsilon(t-1) + \ldots .$$

Then for $h = 1, 2, \ldots$

$$\sigma^2_{h,\infty} = \sigma^2_\infty \{1 + \beta_1^2 + \ldots + \beta_{h-1}^2\} .$$

The graph of $\sigma^2_{h,\infty}$ increases monotonically from σ^2_∞ (at $h = 1$) to 1 (as h tends to ∞). We define [see Parzen (1981)]

$$PVH(h) = 1 - \sigma^2_{h,\infty} , \quad h = 1, 2, \ldots$$

and define horizon HOR to be the smallest value of h for which $PVH(h) \le 0.05$ (or 0.10), whence $\sigma^2_{h,\infty} \ge .95$ (or .90).

To estimate the infinite moving average coefficients we invert the transfer function $g_m(z)$ of an approximating autoregressive scheme to obtain for $k = 1, 2, \ldots$

$$\alpha_0 \beta_k + \alpha_1 \beta_{k-1} + \ldots + \alpha_k \beta_0 = 0 ,$$

The value of the horizon HOR, and the shape of PVH, are used to help classify the type of a time series model; a long horizon indicates a long memory time series.

If the time series type is identified as no memory, the analysis could be considered completed. However, the autoregressive spectral estimation corresponding to the second best order $\hat{m}(2)$, and the log spectral estimator, are examined with the aim of improving the scientific fit rather than the statistical fit.

If the time series type is identified as long memory, or has long memory components as indicated by the prediction variance horizon (PVH) curve, one seeks transformations to a short memory time series $\tilde{Y}(t)$.

STEP V. ARMA Model Identification. If the time series type is identified as short memory one seeks to determine the orders and coefficients of the ARMA whitening filter which transforms the time series to white noise. Quantitative judgements as to whether an MA, AR, or ARMA model is needed are provided by S-PLAY (a display of the S-array of Gray et al (1978) (1980)) and subset ARMA (subset regression selection of ARMA).

STEP V. ARMA Parameter Estimation. The foregoing steps are concerned with model identification. When fitting a model to data, it is customary to test goodness of fit after parameter estimation. In time series analysis, one may be able to test goodness of fit of a model type before estimating its parameters. When one has identified the model for a time series to be stationary and ARMA(p,q), the final step is efficient estimation of the parameters of the ARMA(p,q) scheme.

6. Non-stationary Best Lag Autoregression

The determination of a most appropriate "gentle" transformation of Y to $\tilde{Y}$, where Y is long memory and $\tilde{Y}$ is short memory must inevitably involve the physical nature of the observed time series. Semi-automatic approach can be developed by considering the following examples of long memory time series.

A time series $Y(t)$, $t = 0, \pm 1, \ldots$, is called periodic, with period τ, if

$$Y(t+\tau) - Y(t) = 0, \quad \text{all } t.$$

It follows a linear trend $Y(t) = a+bt$, if for all t

$$Y(t+1) - Y(t) = b, \quad \text{a constant}.$$

It is a pure harmonic of period τ if for all t

$$Y(t) - \phi Y(t-1) + Y(t-2) = 0, \quad \phi = 2\cos\frac{2\pi}{\tau}.$$

Then

$$Y(t) = A\cos\frac{2\pi}{\tau}t + B\sin\frac{2\pi}{\tau}t.$$

As gentle memory shortening transformations, it is natural to consider

$$\tilde{Y}(t) = Y(t) - \hat{\phi}(\hat{\tau})\,Y(t-\hat{\tau}), \tag{1}$$

$$\tilde{Y}(t) = Y(t) - \hat{\phi}_1 Y(t-1) - \hat{\phi}_2 Y(t-2) \tag{2}$$

$$\tilde{Y}(t) = Y(t) - \hat{\phi}_1 Y(t-(m-1)) - \hat{\phi}_2 Y(t-m) \tag{3}$$

whose coefficients τ, $\hat{\phi}(\tau)$, $\hat{\phi}_1$, $\hat{\phi}_2$ are determined adaptively from the data. Our first choice is (1); the lag $\hat{\tau}$ is chosen to minimize over τ

$$Err(\tau) = \sum_{t=\tau+1}^{T} \{Y(t) - \hat{\phi}(\tau)\,Y(t-\tau)\}^2 \div \sum_{t=\tau+1}^{T} Y^2(t)$$

and $\hat{\phi}(\tau)$ is chosen to minimize over $\phi(\tau)$

$$\sum_{t=\tau+1}^{T} \{Y(t) - \phi(\tau)\, Y(t-\tau)\}^2$$

The stationary correlation function $\hat{\rho}(\tau)$ of $\{Y(t),$ $t = 1, 2, \ldots, T\}$ is defined by

$$\hat{\rho}(\tau) = \sum_{t=1}^{T-\tau} Y(t)\, Y(t+\tau) \div \sum_{t=1}^{T} Y^2(t) .$$

Define

$$SSQ(\nu) = \sum_{t=1}^{\nu} Y^2(t) .$$

One can show that

$$\hat{\phi}(\tau) = \hat{\rho}(\tau) \frac{SSQ(T)}{SSQ(T-\tau)}$$

$$Err(\tau) = 1 - |\hat{\phi}(\tau)|^2 \frac{SSQ(T-\tau)}{SSQ(T) - SSQ(\tau)}$$

The most significant lag $\hat{\tau}$ is defined as the value τ minimizing $Err(\tau)$.

It should be noted that $\hat{\rho}(\tau)$ can be easily computed for all values of τ, and therefore so can $\hat{\phi}(\tau)$. One of the aims of our approach is to detect important long lags. As an example, in monthly sunspot series one finds a significant lag to be 124.

We propose three possible actions at the initial stage of analysis of a time series $Y(t)$, $t = 1, \ldots, T$:

L. Declare time series to be long memory, and form $\tilde{Y}(t)$ by (1)

M. Declare time series to be moderately long memory, and form $\tilde{Y}(t)$ by (2) .

S. Declare time series to be short memory, and form $\tilde{Y}(t) = Y(t)$, or $\tilde{Y}(t) = Y(t) - \bar{Y}$

where $\bar{Y}$ is the sample mean. After computing $\bar{Y}$, one performs a naive test to decide if it should be set equal to 0: a naive test is $|\bar{Y}| \leq 2\sigma/\sqrt{T}$ where σ is the sample standard deviation.

1. Compute and print $\hat{\phi}(\tau)$ and $Err(\tau)$ for $\tau = 1, 2, \ldots, M_\tau$, where M_τ is suitably chosen (15 for yearly, quarterly, or monthly data):

2. Determine $\hat{\tau}$. If $Err(\hat{\tau}) \leq 8/T$, go to L.

3. If $\hat{\phi}(\hat{\tau}) \geq .9$, and $\hat{\tau} > 2$, go to L.

4. If $\hat{\phi}(\hat{\tau}) \geq .9$ and $\hat{\tau} = 1$ or 2 determine the best fitting non-stationary AR(2) scheme minimizing

$$\sum_{t=3}^{T} \{Y(t) - \phi_1 Y(t-1) - \phi_2 Y(t-2)\}^2$$

Let $\hat{\phi}_1, \hat{\phi}_2$ denote the minimizing values of ϕ_1 and ϕ_2. Then go to M.

5. If $\hat{\phi}(\hat{\tau}) \leq .9$ go to S.

6. If $\hat{\phi}(\tau)$ is approximately 1 for some τ, one may set this value of τ equal to $\hat{\tau}$ and go to L. One compares the stationary analysis of this choice of memory shortening transformation with that determined by the value of τ minimizing $Err(\tau)$.

7. Non-stationary prediction analysis of a time series in general finds coefficients $\phi_1, \ldots, \phi_m$ minimizing (for a specified memory m)

$$\sum_{t=m+1}^{T} \{Y(t) - \phi_1 Y(t-1) - \ldots - \phi_m Y(t-m)\}^2 .$$

We recommend a <u>subset regression</u> solution which attempts to determine the most significant lags $j_1, \ldots, j_n$ minimizing

$$\sum_{t=m+1}^{T} \{Y(t) - \phi_{j_1} Y(t-j_1) - \ldots - \phi_{j_n} Y(t-j_n)\}^2$$

and determines the solution for a specified set of lags $j_1, \ldots, j_n$. One may take $n = 2$, and j_1 and j_2 are two adjacent lags (m-1 and m) for which $\phi(\tau)$ is approximately 1; one then obtains the transformation of type (3).

A useful guide for selecting alternative models for comparison is the concept of a "second best non-stationary lag" $\hat{\tau}$ at which the lowest relative minimum of $Err(\tau)$ is attained which is greater than the absolute minimum. One may define $\hat{\tau}$ as the lowest index τ such that $Err(\tau-1) > Err(\tau) < Err(\tau+1)$ but $Err(\tau) >$ global minimum of $Err(\tau)$. One should study the ARARMA models whose initial non-stationary autoregression corresponds to the second best lag $\hat{\tau}$.

7. <u>Computer Programs</u>

Our short form forecasting model analysis computer program (called DTFORE) provides the following diagnostics about a time series sample $\{Y(t),\ t = 1, 2, \ldots, T\}$.

1. Name and length of Y.
2. $R(0) = T^{-1} \sum_{t=1}^{T} Y^2(t)$ and $\hat{\rho}(v)$, $v = 1, \ldots, M$, for Y
3. $\hat{\phi}(\tau)$ and $Err(\tau)$, $\tau = 1, \ldots, M$.
4. Non-stationary best lag autoregression coefficients to form $\tilde{Y}(t)$.
5. "Residual variance" $RVY = \sum_{t=\tau+1}^{T} \tilde{Y}^2(t) \div \sum_{t=\tau+1}^{T} Y^2(t)$

   where τ is the maximum lag used to form $\tilde{Y}(t)$.
6. Mean of $\tilde{Y}(t)$.
7. $R(0) = (T-\tau)^{-1} \sum_{t=\tau+1}^{T} \tilde{Y}^2(t)$ and $\hat{\rho}(v)$ for $\tilde{Y}$.
8. CAT(m).
9. CAT1, CAT2 (Best and second best AR orders), PVH, HOR1, HOR2 (prediction variance horizon curve, and horizons computed by assuming AR schemes of orders CAT1, CAT2 respectively).
10. Coefficients of best order approximating AR scheme for $\tilde{Y}$.
11. "Residual variance" $RVYT = \hat{\sigma}^2_{\hat{m}}$ computed from Yule Walker equations, where m = CAT1.
12. Coefficients of best subset AR scheme for $\tilde{Y}$.
13. Coefficients of best subset ARMA scheme for $\tilde{Y}$.

14. Forecasts $Y^{\mu}(T+h|T)$ of $Y(T+h)$ for $h = 1, 2, \ldots, H$, where H is the horizon specified by the user.

The residual variance RVYT is a measure of $E[\epsilon^2(t)] \div E[\tilde{Y}^2(t)]$, where $\epsilon(t)$ are the innovations or white noise term in an ARMA model for $\tilde{Y}(t)$. The variance of $\epsilon(t)$ measured in units of the mean square of $Y(t)$ is estimated by $\hat{\sigma}_\epsilon^2 = (RVY)(RVYT)$.

Our long form time series analysis program called ARSPID (for autoregressive spectral identification) is described in Parzen (1980). It now includes the features of DTFORE.

To evaluate the performance of a model for forecasting h steps ahead, ARSPID has the ability to compute mean absolute percentage error MAPE(h) and mean square error MSE(h) defined by

$$MAPE(h) = \frac{1}{n} \sum_t \left| \frac{Y^{\nu}(t+h|t)}{Y(t+h)} \right|$$

$$MSE(h) = \frac{1}{n} \sum_t |Y^{\nu}(t+h|t)|^2$$

where n is the number of values of t over which the sum is taken.

I would like to express my gratitude to H. Joseph Newton and Scott W. Anderson for their collaboration in this research and for developing the computer programs.

8. Wolfer Sunspot Numbers 1846-1963

The problem of modeling the annual time series of Wolfer's sunspot data has an extensive literature; a review is given by Woodward and Gray (1978). Different authors use data over different time periods; we fit a model from the data for the years 1846-1963 which is sample of length 118. Let $Y(t)$ denote the time series, and $\hat{Y}(t) = Y(t) - \bar{Y}$ denote the time series of derivations from the mean. Almost all authors have defined their models for $\hat{Y}(t)$ rather than for $Y(t)$. Many authors (Yule, Moran, Box and Jenkins) have obtained an AR(2) model for $\hat{Y}(t)$, similar to

$$\hat{Y}(t) - 1.34\,\hat{Y}(t-1) + .65\,\hat{Y}(t-2) = \epsilon(t) \tag{1}$$

This is a short memory model (horizon HOR = 7)

Woodward and Gray propose a model which we write

$$\begin{aligned} \tilde{Y}(t) &= (I-1.64L + L^2)\,\hat{Y}(t) \\ (I-.29L^4 - .21L^5 - .19L^6)\,\tilde{Y}(t) &= (I - .34L)\epsilon(t) \end{aligned} \tag{2}$$

This is a long memory model in our terminology in which $\tilde{Y}$ is ARMA(6,1). While model (1) and model (2) have similar residual variance (235 and 214 respectively), they differ greatly in two important aspects: (1) estimators of spectral density and (2) long range forecasts.

A new model proposed by our program, with moderate memory, is

$$\begin{aligned} \tilde{Y}(t) &= (I - .482L^{10} - .554\,L^{11})\,Y(t), \quad RVY = .149 \\ (I - 1.009L + .362L^2)\,\tilde{Y}(t) &= \epsilon(t), \quad RVYT = .392 \end{aligned} \tag{3}$$

Best Lag Non-stationary Autoregression

The minimum of $Err(\tau)$ was at $\tau = 1$ with $Err(1) = .137$ and $\phi(1) = .926$. Since $Err(1)$ is not less than $8/118 = .068$, our automatic program calls (since $\tau = 1$) for first considering a best non-stationary AR(2) which in this case is

$$\tilde{Y}(t) = Y(t) - 1.467\, Y(t-1) + .586\, Y(t-2) ,$$

with RVT = .088; stationary analysis of $\tilde{Y}(t)$ yields an AR(11) with RVYT = .72 , HOR = 16. This model has innovations (one step ahead prediction errors) of "standard quality" in the sense that their properties are similar to those from other models. To obtain alternative models we determine the second best lags for which $Err(\tau)$ has a relative minimum, at which usually $\phi(\tau) \approx 1$. We obtain

τ	10	11	12	20	21	22
$\phi(\tau)$	.991	1.001	.910	.923	.995	.993
$Err(\tau)$	.190	.174	.317	.399	.302	.307

Our rules conclude from this result that an alternative model has a best non-stationary autoregression of the form

$$\tilde{Y}(t) = Y(t) - \phi_1 Y(t-10) - \phi_2\, Y(t-11) .$$

It is remarkable that $\tilde{Y}(t)$ has mean 1.53 which can be considered approximately 0.

Stationary Analysis of $\tilde{Y}(t)$, of Model (3)

- Both CAT and AIC determine the order of the best approximating AR scheme to be 2, with horizon of 3. Thus $\tilde{Y}(t)$ is clearly

short memory. Figures 1-6 graph $\tilde{Y}$ and the following 5 functions associated with $\tilde{Y}$: $\hat{\rho}$, AIC, CAT, PVH, $\hat{f}_2(\lambda)$.

Combined Analysis of Y(t)

The spectral density of Y(t) implied by the iterated model is presented in Fig. 7; it has sharp spectral peaks one should expect for a moderate memory time series. Fig. 8 graphs Y(t) and $Y^{\mu}(t)$, the best one-step ahead predictors. The ability of the iterated model to provide long range forecasts is indicated by the following table:

Predictors h Steps Ahead	Average Squared Error (MSE)	Average Absolute Percentage Error (MAPE)
1	251	.57
2	427	.69
3	558	.78
4	585	.81
5	590	.81
6	601	.81
7	619	.83
8	626	.73
9	595	.59
10	551	.56
11	658	.74
12	990	1.25
13	1265	1.51
14	1365	1.56
15	1381	1.56
16	1382	1.57
17	1410	1.59
18	1441	1.60
19	1449	1.44
20	1383	.50
21	1332	.35
22	1575	.49

9. International Air Lines Data

As an illustrative example of time series analysis many authors, including Box and Jenkins (1976) have considered a monthly time series of passenger totals in international air travel from 1949 to 1960; it has length T = 144. The model fitted by Box and Jenkins to the time series Y(t) of log international airline data is

$$\tilde{Y}(t) = (I-L)(I-L^{12})\, Y(t) ,$$
$$\tilde{Y}(t) = (I-\theta_1 L)(I-\theta_{12}L^{12})\, \epsilon(t) ; \qquad (1)$$

model (1) is called an "airline model" and is widely fitted to diverse economic time series. For the log international airline data, $\theta_1 = 0.4$, $\theta_{12} = 0.6$, and $\hat{\sigma}_\epsilon^2 = .0013/30.9 = .00004$.

Parzen (1980) recommends the model

$$\tilde{Y}(t) = (I-L^{12})\, Y(t), \ \mathrm{RVY} = .00059$$
$$g_{13}(L)\, \tilde{Y}(t) = \epsilon(t) , \ \mathrm{RVYT} = .08 \qquad (2)$$
$$g_{13}(L) = I-.66L-.31L^2+.14L^3-.06L^4-.18L^5-.02L^6$$
$$+.11L^7-.05L^8-.10L^9+.10L^{10}+.09L^{11}+.26L^{12}-.30L^{13}$$

Another model is: $\tilde{\tilde{Y}}(t) = \tilde{Y}(t)-\tilde{\mu}$, $\tilde{\mu} = E[\tilde{Y}(t)] = .1198$,

$$(I-.74L)(I+.38L^{12})\, \tilde{\tilde{Y}}(t) = \epsilon(t), \ \mathrm{RVYT} = .39,$$

or preferably

$$(I-.55L-.27L^2)\, \tilde{\tilde{Y}}(t) = (I-.49L^{12})\, \epsilon(t) .$$

The model obtained automatically by the techniques proposed in this paper is: $\tilde{\mu} = E[\tilde{Y}(t)] = .0021$,

$$\tilde{Y}(t) = (I-\phi(12)\, L^{12})\, Y(t) , \ \phi(12) = 1.02 , \ \mathrm{RVY} = .00014 ,$$
$$g_{13}(L)\, \tilde{Y}(t) = \epsilon(t) , \ \mathrm{RVYT} = .335 . \qquad (3)$$
$$g_{13}(L) = I-.55L-.27L^2+.06L^3-.02L^4-.10L^5+.08L^6$$
$$+.08L^7-.03L^8-.16L^9+.15L^{10}+.15L^{11}+.28L^{12}-.31L^{13}$$

A subset ARMA model for $\tilde{Y}(t)$ is

$$(I-.78L)\, \tilde{Y}(t) = (I-.42L^{12})\, \epsilon(t)$$

or preferably

$$(I-.57L-.28L^2)\, \tilde{Y}(t) = (I-.45L^{12})\, \epsilon(t)$$

The subset ARMA models may mislead us into thinking that these models are more similar than they are. The effect of $I - L^{12}$ versus $I - 1.02L^{12}$ is shown by the horizons of the AR(13) model fitted to the residuals $\tilde{Y}(t)$. For the old model (2), HOR = 65 while for the new model (3) HOR = 13. The trend that is not annihilated by pure 12th differencing is successfully modeled by the AR(13) model fitted to the residuals.

Graphs of the fit of $Y^{\mu}(t)$ to Y(t) in Figures 9 and 10 indicate that models (2) and (3) fit equally well. However their average forecast errors, listed in Table B, indicate that model (3) is superior.

It should be emphasized that "airline model" which is adopted by many researchers as a model for economic time series would not in general be regarded as adequate by the criteria we propose, which do not recommend double differencing as a memory shortening transformation. When the need

for double differencing arises, it appears as a situation in which long memory components continue to be present even after several iterations; then the final iterated model is of the form

$$Y(t) \longrightarrow \boxed{} \longrightarrow Y^{(1)}(t) \longrightarrow \boxed{} \longrightarrow Y^{(2)}(t) \longrightarrow \boxed{} \longrightarrow \epsilon(t)$$

which we call a <u>randem</u> filter (randem is a word derived from tandem to describe three horses harnessed tandem).

As a matter of historical interest, randem filter models were discussed and recommended in Parzen (1964),(1967). However it was not realized there that the first filter should be a non-stationary autoregression.

10. <u>Metals Series of Makridakis</u>

A monthly time series of length 144 representing carbon steel monthly shipments from 1961-1972 is discussed by Makridakis (1978). The series has mean μ = 3082, variance 324000, and standard deviation σ = 570. The coefficient of variation (σ/μ) equals 18. It can be regarded as a rough upper bound for average relative prediction error. If an ARIMA model is chosen to fit Y(t), Makridakis finds the model, written in our notation,

$$\begin{aligned} Y(t) &= (I-L)\,Y(t) \\ \tilde{Y}(t) &= (I-.45L)(I+.26L^{12})\,\epsilon(t) \end{aligned} \qquad (1)$$

Parzen (1980) fits the model

$$\begin{aligned} \hat{Y}(t) &= Y(t) - \hat{Y}\ , \quad RVY = .03\ , \\ g_{13}(L)\,\hat{Y}(t) &= \epsilon(t)\ , \quad RVYT = .47 \end{aligned} \qquad (2)$$

$$g_{13}(L) = I-.66L-.31L^2+.14L^3-.06L^4-.18L^5-.02L^6 +.11L^7-.05L^8-.10L^9+.10L^{10}+.09L^{11}+.26L^{12}-.30L^{13}$$

A subset autoregression model for $\hat{Y}(t)$ is

$$(I-.47L-.30L^2-.3L^{12}+.22L^{13})\,\hat{Y}(t) = \epsilon(t)\ .$$

An ARARMA model fitted automatically for Y(t) is

$$\begin{aligned} \tilde{Y}(t) &= (I-.99L)\,Y(t)\ , \quad RVY = .021\ , \\ g_1(L)\,Y(t) &= \epsilon(t),\ RVYT=.79,\ g_1(L) = I + .463L \end{aligned} \qquad (3)$$

The best model for $\tilde{Y}(t)$ is an AR(13), since the CAT criterion function has almost equal values at lags 1 and 13, CAT2 = 13, and AIC chooses 13. Our subset ARMA program also

chooses AR(13), and does not fit on MA part. A subset autoregression is

$$(I + .411L - .295L^{12})\ Y(t) = \varepsilon(t)\ , \quad RVYT = .70$$

Makridakis reports a paradox about the metals series that the model

$$\tilde{Y}(t) = (I-L)\ Y(t)\ , \quad \tilde{Y}(t) = \varepsilon(t) \text{ white noise} \qquad (4)$$

yields smaller forecast errors than the "optimum" Box-Jenkins model (1). Table C lists MAPE (mean absolute percentage errors) for various forecast horizons and various models. Our model (3) yields smaller forecast errors than Makridakis' models (1) and (4). However even smaller errors are attained by our model (2) or by a model whose theory is described in Parzen and Pagano (1979):

$$\begin{aligned} &Y(t) \text{ is periodically stationary,} \\ &Y(t) - \text{monthly means is } AR(2) \end{aligned} \qquad (5)$$

An important alternative model which may be superior for long range forecasting is provided by, in the initial non-stationary autoregression, choosing the second best lag, corresponding to the lowest relative minimum of Err. The model implied by this approach for the carbon metals series is

$$\begin{aligned} &\tilde{Y}(t) = (I-1.003L^{12})\ Y(t)\ , \quad RVY = .009 \\ &g_2(L)\ Y(t) = \varepsilon(t)\ , \quad RVYT = .48 \\ &g_2(L) = I - .50L - .28L^2 \end{aligned} \qquad (6)$$

The mean average percentage error of forecasting h steps ahead using models (1) to (6) are compared in Table C. Among ARARMA models, (6) seems clearly best on the basis of Table C. An alternative criterion for comparing models is how well their one-step ahead forecasts $Y^{\mu}(t)$ fit $Y(t)$. In Figures 11 and 12, the solid curve are the values of $Y(t)$, and the crosses are $Y^{\mu}(t)$ for model (3) in Fig. 11 and model (6) in Fig. 12. One might prefer model (3) if the comparison is based on the behavior of one-step ahead predictors.

A time series forecaster must evaluate the value of a proposed model in the context of the actual time series with which he, or she, is concerned. Our procedures aim to be applicable in all the diverse fields to which time series analysis is being applied by providing the researcher automatically with several alternative models for consideration which are obtained by combining the best or second best non-stationary autoregression to generate $\tilde{Y}(t)$ with the best or second best stationary autoregression for transforming $\tilde{Y}(t)$ to $\varepsilon(t)$.

It may happen that the transformation yielding $\tilde{Y}(t)$ is a pure differencing operator, and the model belongs to the ARIMA class. We believe it should still be called an ARARMA model, because the model was found by the systematic procedure described in this paper rather than by merely guessing the form of the difference operator and checking the guess by analysis of residuals.

Table A

Non-stationary Best Lag Diagnostics

	Carbon Steel		International Airline Data (log)	
Lag τ	Err(τ)	Phi(τ)	Err(τ)	Phi(τ)
1	.021	.990	.00037	1.001
2	.022	.992	.00088	1.003
3	.028	.990	.00131	1.004
4	.032	.989	.00163	1,006
5	.038	.986	.00169	1.008
6	.040	.988	.00167	1.010
7	.041	.987	.00169	1.012
8	.041	.989	.00164	1.013
9	.042	.990	.00133	1.015
10	.042	.992	.00095	1.017
11	.043	.991	.00046	1.019
12	.037	.994	.00013	1.021
13	.047	.990	.00047	1.023

Table B

Log International Airline Data

Mean Average Percentage Error (MAPE) and

Mean Square Error (MSE) of

Forecasts Horizon h Ahead

Forecast	MAPE (in percent)		MSE	
Horizon	Model (2)	Model (3)	Model (2)	Model (3)
1	.38	.36	.00084	.00083
2	.43	.41	.00102	.00099
3	.46	.43	.00136	.00128
4	.54	.48	.00159	.00156
5	.55	.49	.00178	.00166
6	.60	.53	.00209	.00191
7	.68	.58	.00241	.00217
8	.68	.58	.00258	.00227
9	.73	.59	.00279	.00241
10	.78	.65	.00335	.00281
11	.80	.68	.00353	.00294
12	.81	.69	.00337	.00308
13	.89	.79	.00484	.00381
14	.87	.82	.00501	.00397
15	.98	.88	.00508	.00451
16	1.03	.95	.00675	.00503
17	1.07	.99	.00715	.00529
18	1.15	1.03	.00780	.00564
19	1.18	1.08	.00870	.00612
20	1.23	1.10	.00894	.00629
21	1.24	1.13	.00951	.00655
22	1.28	1.18	.01040	.00696

The reader may find it interesting to compare square root of MSE with estimated standard deviation of forecast errors reported for model (1) by Box and Jenkins (1976), p. 311. Note that MSE, unlike $\hat{\sigma}_\epsilon^2$, is not measured in units of $E[Y^2(t)]$.

Table C

Carbon Metals Series
Mean Average Percentage Errors (MAPE)
of Different Models when Forecasting h Steps Ahead

Forecast	Makridakis		Parzen			
Horizon	(4)	(1)	(2)	(5)	(6)	(3)
1	10.88	11.42	8.65	7.63	9.79	9.89
2	13.55	13.21	9.86	8.03	10.80	11.41
3	15.87	14.87	11.04	9.21	11.45	12.99
4	18.05	15.22	11.76	10.09	12.19	13.86
5	18.77	15.66	12.20	10.54	12.81	14.89
6	18.84	16.31	12.31	10.63	12.98	15.18
7			12.45	11.00	13.30	16.38
8			12.48	11.41	13.44	16.60
9	18.60	17.03	12.40	11.51	13.97	16.81
10			12.30	11.70	13.83	16.03
11			12.17	11.83	14.09	15.77
12	18.45	18.49	12.04	11.82	14.41	16.05
13					16.66	16.60
14					18.00	17.34
15					18.62	18.27
16					19.20	19.49
17					19.15	20.07
18					19.24	21.18
19					19.77	21.60
20					20.07	21.47
21					20.28	21.34
22					20.47	20.54

11. Time Series Decomposition and Seasonal Adjustment

One of the traditional aims of economic time series analysis has been time series decomposition, which can be defined as a representation of a time series $Y(t)$ as the sum of a trend component $P(t)$, a seasonal component $S(t)$, and an irregular component $Z(t)$:

$$Y(t) = P(t) + S(t) + Z(t) .$$

Let τ denote the seasonal period (12 for monthly time series). *An approach to forming $P(t)$ is to pass $Y(t)$* through a low pass filter which suppresses components of period τ. A possible definition is

$$P(t) = \sum_{j=0}^{2\tau} Y(t-j)\, w(1-\tfrac{j}{\tau})$$

where we choose w to be the Bohman kernel

$$w(u) = (1-|u|) \cos \pi u + \frac{1}{\pi} \sin \pi u ;$$

one could also choose $w(u)$ to be the Parzen kernel [Parzen (1967)].

The time series $P(t)$ would be a long memory time series one would model in order to form forecasts $P(t+h|t)$. Similarly one would model

$$\tilde{Y}(t) = Y(t) - P(t)$$

in order to form forecasts $\tilde{Y}(t+h|t)$. One would expect to be able to decompose

$$\tilde{Y}(t) = S(t) + Z(t) ,$$

perhaps by treating $\tilde{Y}(t)$ as a periodically stationary time series. One hopes to not only form forecasts $Y(t+h|t)$ but also a time series decomposition. Research on this proposal is in progress.

References

Akaike, H. (1976) "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion," System Identification, edited by R.K. Mehra and D.G. Lainiotis, Academic Press: New York, 27-96.

Akaike, H. (1979) "A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Fitting" Biometrika, 66, 237-242.

Box, G.E.P. and Jenkins, G.M. (1976) Time Series Analysis, Forecasting, and Control revised edition, Holden Day: San Francisco.

Gray, H.L., Kelley, G.D., and McIntire, D.D. (1978) "A New Approach to ARMA Modeling", Communications in Statistics, B7, 1-77.

Hannan, E.J. and Quinn, B.G. (1979) "The Determination of the Order of an Autoregression", Journal of the Royal Statistical Society, 41, 190-195.

Makridakis, S. (1978) "Time Series Analysis and Forecasting: an Update and Evaluation", *Int. Statist. Rev.*, 46, 255-273.

Makridakis, S. and Hibon, M. (1979) "Accuracy of Forecasting: An Empirical Investigation", *Journal Royal Statist. Soc. Series A*, 142, 97-145.

Parzen, E. (1964) "An Approach to Empirical Time Series", J. Res. Nat. Bur. Standards, 68D, 937-951.

Parzen, E. (1967) Time Series Analysis Papers, Holden Day: San Francisco.

Parzen, E. (1967) "Empirical multiple time series analysis" Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, edited by L. LeCam, and J. Neyman, University of California Press, Vol. I, pp. 305-340.

Parzen E. (1967) "The Role of spectral analysis in time series analysis" Review of the International Statistical Institute, Vol. 35, (1967), 125-141.

Parzen, E. (1974) "Some Recent Advances in Time Series Modeling" IEEE Transactions on Automatic Control, Vol. AC-19, No. 6, December 1974, 723-730.

Parzen, E. (1976) "An Approach to Time Series Modeling and Forecasting Illustrated by Hourly Electricity Demands" Technical Report, Statistical Science Division, SUNY at Buffalo.

Parzen, E. (1977) "Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes" Multivariate Analysis - IV, edited by P. Krishnaiah, North Holland: Amsterdam, 283-295.

Parzen, E. and Pagano, M. (1979) "An Approach to Modeling Seasonally Stationary Time Series", Journal of Econometrics, 9, 137-153.

Parzen, E. (1979) "Forecasting and Whitening Filter Estimation" TIMS Studies in the Management Sciences, 12 (1979) 149-165.

Parzen E. and Newton, H.J. (1980) "Multiple Time series modeling, II" Multivariate Analysis - V, edited by P. Krishnaiah, North Holland: Amsterdam. 181-197.

Parzen, E. (1980) "Time Series Modeling, Spectral Analysis, and Forecasting" Directions in Time Series Analysis, ed. D.R. Brillinger and G. C. Tiao, Institute of Mathematical Statistics.

Parzen, E. (1981) "Modern Empirical Statistical Spectral Analysis". *Journal of Time Series Analysis*, Vol. 1, No. 1.

Parzen, E. (1981) "Time Series Model Identification and Prediction Variance Horizon", Second Tulsa Symposium on Applied Time Series Analysis, ed. David Findley, Academic Press: New York.

Woodward, W.A. and Gray, H.L. (1978) "New ARMA Models for Wolfer's Sunspot Data", *Commun. Statist. Simulation and Computation*, *B7*, 97-115.

Woodward, W.A. and Gray, H.L. (1980) "On the Relationship between the R and S Arrays and the Box-Jenkins Method of ARMA Model Identification",

Fig. 1

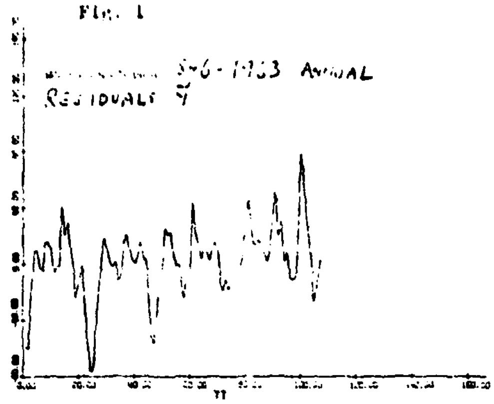


Fig. 3

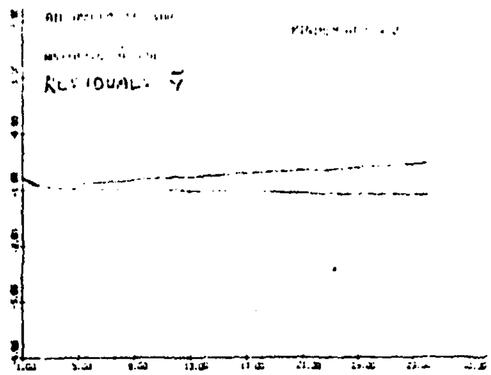


Fig. 2

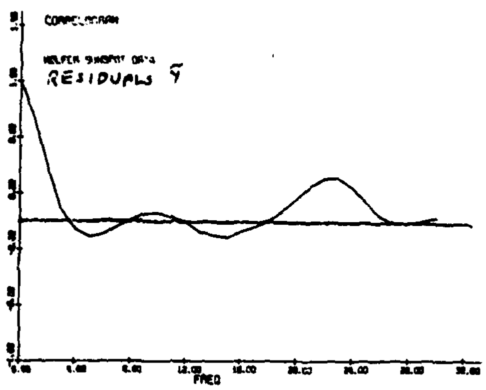


Fig. 4

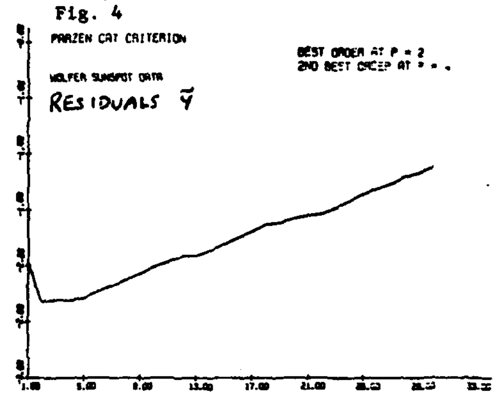


Fig. 5

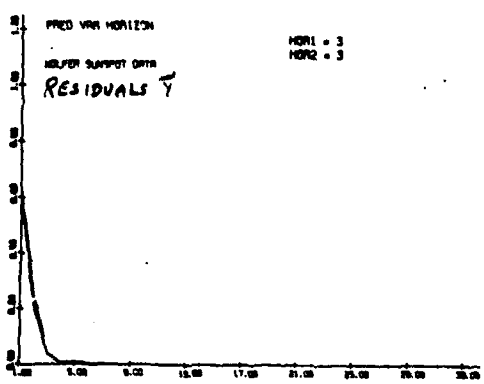


Fig. 7

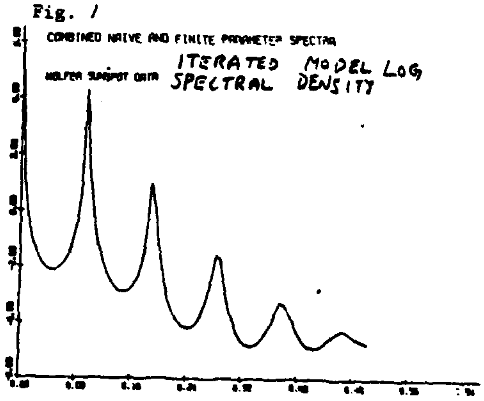


Fig. 6

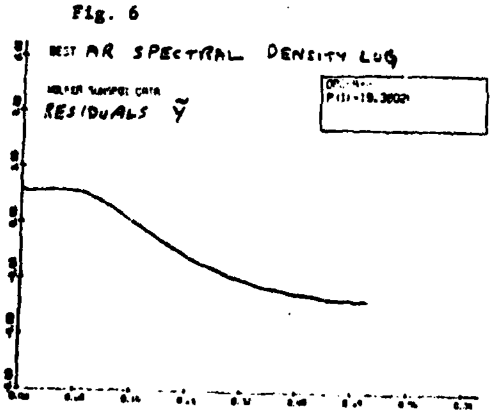


Fig. 8

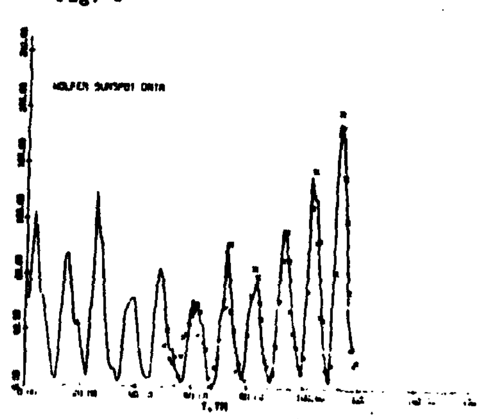


Fig. 9

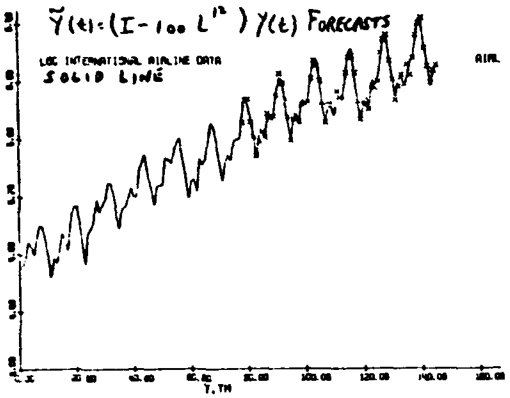


Fig. 10

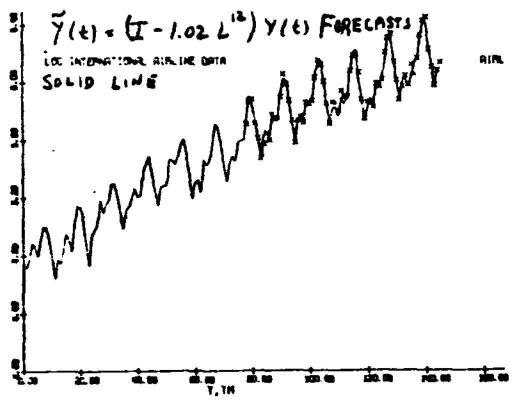


Fig. 11

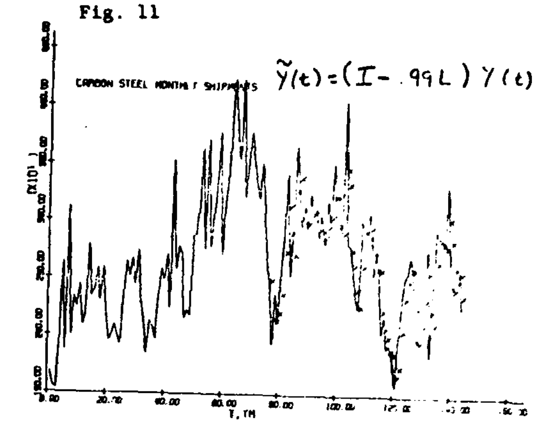
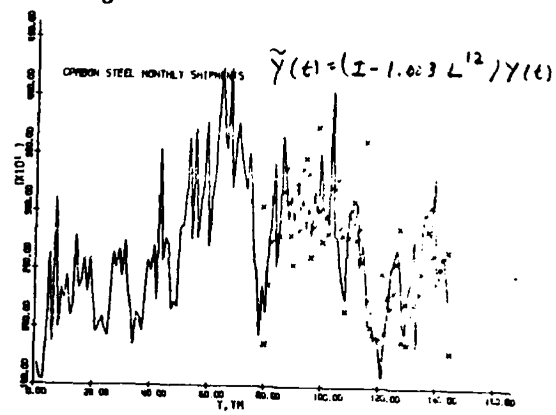


Fig. 12



8